Nucleic Acids Research, Vol. 19, No. 24 6763–6769

Tomato yellow leaf curl virus from Sardinia is a whitefly-transmitted monopartite geminivirus

Ahmed Kheyr-Pour[1], Mohammed Bendahmane[1], Volker Matzeit[2], Gian Paolo Accotto[3], Stefania Crespi[3] and Bruno Gronenborn[1,2]*

[1]Institut des Sciences Végétales, CNRS, 91198 Gif-sur-Yvette Cedex, France, [2]Max-Planck-Institut für Züchtungsforschung, Carl Von Linné Weg 10, D-5000 Köln 30, FRG and [3]Istituto di Fitovirologia Applicata, CNR, Strada delle Cacce 73, 10135 Torino, Italy

Received October 7, 1991; Revised and Accepted November 28, 1991 EMBL accession no. X61153

ABSTRACT

The genome of an isolate of tomato yellow leaf curl virus from Sardinia, Italy (TYLCV-S), a geminivirus transmitted by the whitefly *Bemisia tabaci*, has been cloned and sequenced. The single circular DNA molecule comprises 2770 nucleotides. Genome organisation closely resembles that of the DNA A component of the whitefly-transmitted geminiviruses with a bipartite genome. A 1.8 mer of the TYLCV-S genome in a binary vector of *Agrobacterium tumefaciens* is infectious upon agroinoculation of tomato plants. Typical tomato yellow leaf curl disease symptoms developed about three weeks after inoculation. The disease was transmitted by the natural vector *B.tabaci* from agroinfected plants to test plants, reproducing in this way the full biological cycle and proving that the genome of TYLCV-S consists of only one circular single-stranded DNA molecule. Contrary to the other whitefly-transmitted geminiviruses described so far, there is no evidence for the existence nor the necessity of a second component (B DNA) in the TYLCV-S genome.

INTRODUCTION

Geminiviruses are a group of small plant viruses with a circular single-stranded DNA genome (1), encapsidated within characteristic twinned isometric particles (2). A detailed description of the important features of a variety of geminiviruses, including genome organization, insect vector specificity and host-virus relationships can be found in reviews by Stanley, (3) and Davies and Stanley, (4).

Geminiviruses transmitted by leafhoppers possess a monopartite genome, and those transmitted by the whitefly *Bemisia tabaci*, until present, appeared to possess a bipartite genome. The two genome components are designated DNA A and B (or DNA 1 and 2). DNA A on its own is capable of replication using the host machinery of DNA synthesis (5,6), whereas DNA B requires in addition the expression of the AL1 or C1 protein encoded by DNA A for its replication (7). In turn, the expression of proteins encoded by DNA B (DNA 2) is necessary for the systemic movement of both DNAs within the plant host and for the elicitation of disease symptoms.

Tomato yellow leaf curl is one of the most devastating virus diseases of cultivated tomato (*Lycopersicum esculentum* Mill.) (8,9). It is transmitted by the whitefly *Bemisia tabaci* (Gennadius), and its causative agent, tomato yellow leaf curl geminivirus (TYLCV), has been identified (10). The disease is endemic in the Eastern Mediterranean basin, in some subtropical African countries and in Latin America (11). A similar disease in Mexico, 'chino del tomate', is also caused by a geminivirus (12).

In 1988 and 1989, severe outbreaks of tomato yellow leaf curl disease occurred for the first time in the Northern part of the Mediterranean basin, in Sardinia and Sicily, (13,14). Symptoms of infected tomato plants from different geographical areas seem to be similar, however, the molecular relationships between the tomato yellow leaf curl viruses responsible for the disease were as yet not clear.

Here we report the DNA sequence and the genome organization of a TYLCV isolate from Sardinia (TYLCV-S). Its cloned single DNA component is sufficient to elicit the typical symptoms of yellow leaf curl disease of tomato upon agroinoculation. In addition, *Bemisia tabaci* transmits the disease from the agroinoculated plants to test tomatoes. This proves that the genome of TYLCV-S consists of only a single DNA component. These findings contrast with those of a recent report by Rochester *et al.*, (15) who describe a TYLCV isolate from Thailand (TYLCV-T) with a bipartite genome.

MATERIALS AND METHODS

Identification and cloning of TYLCV-S

In December 1988 plants showing yellow leaf curl symptoms similar to those described by Cohen and Nitzany, (9) were collected from a tomato field in Sardinia. The disease could easily be transmitted by grafting to tomato (*Lycopersicum esculentum* cv Marmande) and *Datura stramonium* L. Total DNA was

* To whom correspondence should be addressed at Institut des Sciences Végétales, CNRS, 91198 Gif-sur-Yvette Cedex, France

extracted from infected tomato plants by grinding aliquots of tissue frozen in liquid N_2 in extraction buffer (100mM Tris pH 8.0, 100mM NaCl, 50mM EDTA, 0.5% SDS; 3 ml/g leaf), followed by several extractions with phenol and phenol/chloroform. The DNA was precipitated by ethanol and analysed in Southern blot experiments using as probe a 2.7 kb fragment derived from plasmid pTYH 20.7 containing a full length clone of the Israeli isolate of TYLCV (16), labelled by nick translation (17). The hybridisation conditions were: 0.9 M NaCl, 50mM NaH_2PO_4, 5mM EDTA, 0.1% SDS pH 7.0, BSA, Ficoll, PVP each 0.2%, 500mg/ml sonicated and denatured salmon sperm DNA, 50% formamide at 42°C. The filters were washed three times under non stringent conditions: 3×SSC/0.5% SDS, 60°C. Enzymes that only linearized the circular double-stranded TYLCV-S DNA (about 2.7kbp) were used to clone it in the bacterial plasmid pUC 118 (18). Double- and single-stranded DNA of recombinant plasmids was prepared from *E.coli* strain DH5a. Standard techniques in molecular biology were applied as described (19). Two independent clones containing inserts of about 2.7 kbp (pTY *Sst* 14 and pTY *Sph* 14) were further analyzed and used to determine the complete DNA sequence of TYLCV-S.

Determination of the DNA sequence

For sequencing the dideoxynucleotide chain termination method was employed (20). Since the two cloning sites used (*Sst* I and *Sph* I) are located at either end of the polylinker of plasmid pUC118, a set of nested deletions could easily be produced by linearizing the bacterial vector at the single *Sma* I site of the polylinker, cutting the TYLCV insert at convenient (mostly unique) restriction sites, followed by treatment with Klenow fragment of polymerase I in order to create blunt ends and religation. Since the two independent TYLCV-specific inserts in the plasmids pTY-*Sst*14 and pTY-*Sph*14 have the same orientation, the two sets of deletions extend progressively from both sides of the polylinker into the cloned TYLCV DNA. Applying this strategy, no subcloning of fragments was required to position a maximum of different DNA sequences into the reach of the *lac* sequencing primer (in plasmid pTY-*Sst*14) and the *lac* reverse sequencing primer (in plasmid pTY-*Sph*14). Phage M13K07 was used as helper to produce single-stranded DNA. The complete sequence of the TYLCV plus strand was determined from clone pTY-*Sst*14 on single-stranded DNA, whereas the complete sequence of the minus strand was determined from clone pTY-*Sph*14 by reverse priming on double-stranded DNA. To sequence a few regions that could not be covered by this strategy, specific oligonucleotide primers were synthesized (380B, Applied Biosystems).

In this way the complete DNA sequence of two independent TYLCV clones was determined on their respective complementary strands, Figure1.

Construction of clones for agroinoculation

Since TYLCV or its cloned DNA are not transmitted mechanically, we used the agroinoculation technique (21) to assay the infectivity of the cloned DNA. The TYLCV genome was inserted into the *Sst* I site of the binary plant transformation vector pBin19 (22), and a *Bam* HI-fragment of about 600bp extending from the polylinker of pBin19 through the intergenic region of TYLCV to the *Bam* HI site at map position 152 of TYLCV was deleted. Subsequently a complete genome unit of TYLCV was inserted into the remaining *Sst* I site yielding a 1.8 mer of the TYLCV-genome in pBin19 (pBin19/TYLCV-S1.8). This plasmid was propagated in *E. coli* strain DH5a and introduced into *Agrobacterium tumefaciens* strain LBA4404, harbouring the pAL4404 Ti-plasmid following the protocol of Höfgen and Willmitzer (23). pAL4404 is a deletion mutant of pTiAch5T lacking the T-region (24,25). Transformants were selected on YEB plates containing rifampicin (150mg/ml), kanamycin (100mg/ml) and neomycin (20mg/ml) and purified by restreaking on minimal sucrose plates containing the same antibiotics. The pBin19/TYLCV-S1.8 was also transferred into the *A.tumefaciens* strain C58/CI, cured of its Ti-plasmid pTiC58. This strain was used as a negative control in the agroinoculation assays. The presence in *A.tumefaciens* of the authenthic pBin19/TYLCV-S1.8 plasmid was verified by a modified alkaline minipreparation of plasmid DNA adapted to *Agrobacterium.*

Agroinoculation and analysis of plants infected with the TYLCV-S genome

Two tomato cultivars, *L.esculentum* cv Monique (provided by J.C. Mercier, Clause Co., Brettigny sur Orge) and cv Mecline (provided by H. Laterrot, INRA, Avignon) were used for agroinoculation. Young plants were inoculated either at the 3–4 leaf stage or at the 8–10 leaf stage. *A.tumefaciens* cultures were grown at 28°C for about 48 hours, the cells were pelleted and washed twice with water and resuspended in 1/10th of the initial volume of sterile water. To inoculate younger plants a 18-gauge needle was used to inject the concentrated *Agrobacteria* into the petioles of the three youngest leaves. For older plants, the *Agrobacteria* were injected either into the petioles of the youngest leaves or into the decapitated main stem. The plants were placed in a closed growth chamber with 16h-day light at 24°C/70% relative humidity.

Three independent agroinoculation assays were performed. In the first assay, a total of 20 tomato plants were agroinoculated. Ten plants each of the two cultivars Monique and Mecline were inoculated with *Agrobacterium* strains LBA4404/pBin19/TYLCV-S1.8 or with C58/pBin19/TYLCV-S1.8, yielding lots of five plants for each combination. In a second experiment only LBA4404/pBin19/TYLCV-S1.8 was used to inoculate 8 plants each of the two tomato cultivars. The third assay comprised of five tomato plants (Monique) and four tobacco plants (*Nicotiana tabacum*, var. *Xanthi*) inoculated with LBA4404/pBin19/TYLCV-S1.8.

Detection of viral DNA forms in agroinoculated plants

Two methods were used to detect TYLCV DNA in plants. The presence of single-stranded viral DNA in agroinfected plants was assayed using the leaf squash technique described (16). In addition, total DNA was prepared from the plants as in (26). Restriction enzyme digests of the DNA were carried out as recommended by the suppliers and the products were fractionated on 0.8% agarose gels. The DNA was transferred to Nylon membranes (Amersham Hybond) by vacuum or capillary blotting, and the membranes were hybridized to TYLCV-specific probes labelled with ^{32}P by the multiprime technique (27). The membranes were finally washed stringently at 70°C, 0.1×SSC,

Transmission of TYLCV by *Bemisia tabaci*

All experiments involving *B.tabaci* were carried out under conditions to guarantee containment of the insects. Two tomato plants displaying clear symptoms of TYLCV disease (of both Monique and Mecline cultivars) were exposed to several adult

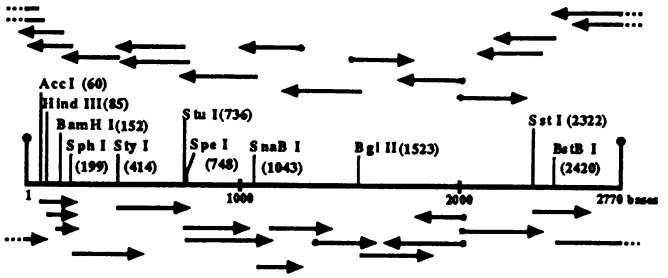

Fig. 1. Nucleotide sequencing of TYLCV DNA. Two independent full-length clones of TYLCV-S (pTY-*Sst*14 and pTY-*Sph*14) were used to produce a set of nested deletions using the restriction enzymes indicated. Arrows above the map represent the extent of sequence information obtained from pTY-*Sst* 14, arrows below the map represent sequence information from pTY-*Sph* 14, knobbed arrows represent sequences obtained using synthetic oligonucleotides at the respective positions. Nucleotide 1 of the map is the initial *T* of the conserved nonanucleotide *TAATATTAC*. The sequence has been deposited at EMBL/GenBank under accession No. X61153.

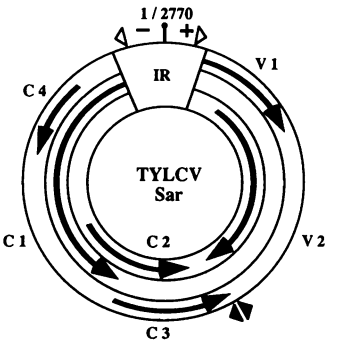

Fig. 2. Genome organisation of TYLCV-S. Open reading frames (ORFs) are shown as black arrows. The ORFs of plus-strand (virion strand) polarity are designated V1 and V2, ORFs of minus-strand (complementary strand) are designated C1 through C4. IR indicates the intergenic region, the position of the stemloop is marked by (|). Open triangles indicate the positions of eukaryotic promoter signals (CAAT and TATAA boxes) in both (+) and (−) sense orientation. Filled triangles indicate the positions of polyadenylation signals (AATAAA), located at the end of ORFs V_2 and C_3.

insects in a closed cage. After an acquisition period of 3 days the insects were transferred for another 3 days to uninfected young tomato seedlings (*L. esculentum* cv. Marmande) having 2 to 4 true leaves. After exposure, the *Bemisia* were killed by repeated spraying with a 1/1 mixture of the insecticides Decis (Sipcam, Milano) and Orthene (Roussel-Hoechst-Agrovet, Milano), and the plants were grown under insect-proof conditions. The appearance of disease symptoms was monitored and the presence of TYLCV was verified by squash blots and Southern blots as described above.

RESULTS

Organization of the TYLCV genome

Southern blot analysis following restriction with several enzymes (e.g. *Bam* H1, *Bgl* II, *Eco* RI, *Hpa* II, *Sca* I, *Sph* I and *Sst* I) had already revealed a considerable restriction site polymorphism between the Israeli isolate of TYLCV and the Sardinian one. Two independent clones (pTY-*Sst*14 and pTY-*Sph*14) were sequenced as outlined in Fig. 1, and the complete DNA sequence of them on both strands was obtained. It comprises 2770 base pairs; (EMBL/GenBank accession No. X61153). The first nucleotide (*T*) of the absolutely conserved nonanucleotide *TAATATTAC* of the stemloop structure occurring in the genome of all geminiviruses (28) is defined as base No 1 of the circular genome of TYLCV-S.

The coding capacity of the TYLCV-S genome is illustrated in Figure 2: In the virion sense orientation (plus-strand) two potential proteins (V_1:13kD and V_2: 30kD) are encoded. In the complementary sense orientation (minus-strand) four open reading frames are found. (C_1 to C_4, Fig.2).

Table 1 summarizes the coordinates of the ORFs and the sizes of their potential proteins; only ORFs with a coding capacity of more than 10kD are displayed. The existence of ORFs on both DNA strands of the genome requires the transcription of both strands for expression. In the intergenic region eukaryotic promoter signals (*TATAA* and *CAAT* boxes) oriented in both senses (Fig. 2, open triangles) are located. Polyadenylation signals (AATAAA, Fig. 2 filled triangles) are located at the end of ORFs V_2 and C_3, respectively. By analogy with other whitefly-transmitted geminiviruses, V2 is assigned to encode the viral capsid protein. It is preceded and partially overlapped by a small ORF (V1) with a coding capacity of 13 kD. An ORF of similar size to V1 is also found in the genome of beet curly top virus (BCTV), (29) and on the DNA A of African cassava mosaic virus (ACMV) (30) as well as on the genomes of all the leafhopper-transmitted geminiviruses. The ORFs of (−) sense orientation, designated C_1(41kD), C_2 (16kD), C_3 (16kD) and C_4 (11.5kD) have their counterparts on DNA A of the geminiviruses infecting dicotyledonous plants (3,4).

Table 1. Open reading frames (ORF) of TYLCV-S

RF	polarity	nucleotide	a. a.	mol. wt.
V1	(+)	146−491	115	13,242
V2	(+)	307−1075	256	29,779
C1	(−)	2614−1537	359	40,709
C2	(−)	1626−1221	135	15,709
C3	(−)	1478−1076	134	15,944
C4	(−)	2463−2169	98	11,461

Table 1. summarizes the nucleotide start and stop coordinates of the open reading frames (ORFs), the number of amino acids (a.a.) and the molecular weight (mol. wt.; daltons) of the potential proteins encoded by TYLCV-S.

Symptom appearance and analysis of viral DNA forms in agroinoculated plants

The characteristic symptoms of disease, yellowing and curling of the youngest leaves, appeared between 14 and 21 days after agroinoculation. At 21 days after inoculation a small piece of a young leaf from each plant was squashed onto a Nylon filter, and without denaturation, the membrane was hybridised with a ^{32}P-radiolabelled TYLCV-specific probe. All 10 plants agroinoculated with the *A. tumefaciens* strain LBA4404 pBin19/TYLCV-S1.8 showed strong hybridisation signals. Squashes of leaves from naturally infected plants were used as positive controls, whereas all 10 plants that had been agroinfected with the *A. tumefaciens* strain C58/C1 pBin19/TYLCV-S1.8 cured of its Ti-plasmid did not exhibit any hybridisation signal. This proved that the positive hybridisation was due to the presence

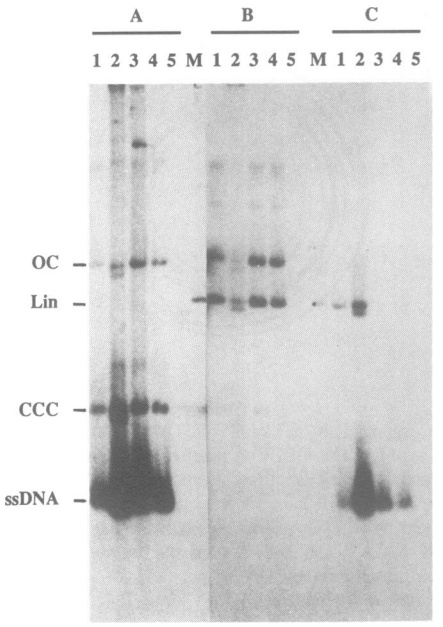

Fig. 3. Forms of TYLCV DNA in naturally infected and agroinfected tomato plants. Southern blot analysis of DNA isolated from TYLCV-infected tomato leaves. **A.** native viral DNA forms, **B.** digested with S1-nuclease, **C.** digested with *Sst* I. Lane 1: DNA of a naturally infected plant from Sicily. Lane 2: DNA of a naturally infected plant from Sardinia. Lanes 3 and 4: DNA of plants agroinoculated with TYLCV-S, Lane 5: DNA of a healthy plant. M: linear genome of TYLCV-S released from plasmid pTY*Sst*14 by cleavage with *Sst* I. The forms of the double-stranded DNA are indicated: open circular (OC), linear (Lin) and covalently closed circular (CCC), single-stranded DNA is marked (ssDNA). The faster migrating DNA in lanes A2, B2 and C2 (field isolate from Sardinia) probably represents a deletion mutant. Gel purified full length TYLCV DNA was used as a probe.

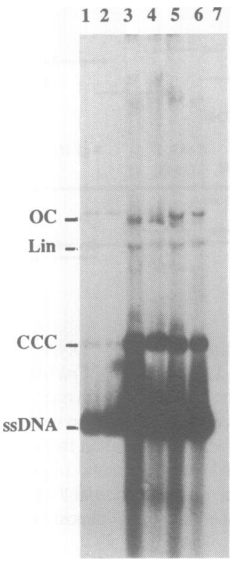

Fig. 4. Forms of TYLCV DNA following transmission by *Bemisia tabaci*. Southern blot analysis of DNA isolated from plants agroinoculated with TYLCV-S (Lanes 1 and 2) and plants infected by *B.tabaci* fed on agroinoculated plants (Lanes 3 to 6). Lane 7: DNA of a healthy plant. The DNA forms of TYLCV are labelled as in figure 3.

of free single-stranded TYLCV DNA in the plant rather than to residual *Agrobacterium* DNA.

Total DNA extracted from the diseased plants and analysed in Southern type experiments revealed the presence of the same structural forms of the TYLCV DNA that had been detected in the DNA from naturally infected plants from Italy. Two forms of single-stranded DNA (circular and linear) and the three forms of the double-stranded replicative intermediates, covalently closed circles (ccc), open circles (oc) and linear (lin), were readily detected (Figure 3A). Treatment with S1 nuclease resulted in the complete digestion of the single-stranded DNA forms. The covalently closed circular forms were converted into open circles by nicking, and a fraction of the open circular forms was linearised by the S1 nuclease, probably at the position of the nicks (Figure 3B). Digestion of the DNA with a restriction endonuclease that cuts the TYLCV DNA only once (*Sst* I) yields the linear double-stranded DNA form, but does not affect the single-stranded DNA (Figure 3C). In a second experiment a total of 16 plants were infected with *A.tumefaciens* strain LBA 4404 containing the pBin19/TYLCV-S1.8 genome. All of them developed disease symptoms after about three weeks, and all of them gave positive hybridisation signals when tested in Southern blot experiments.

In addition, four tobacco plants (*N. tabacum* , cv. *Xanthi*) that were agroinoculated developed no symptoms of disease during a period of about two months. However, one plant out of four showed strong positive hybridisation signals in squash blots. A detailed study on the distribution of the TYLCV DNA within this plant showed that ss-DNA of TYLCV was present throughout the plant, with a particularly high abundance in the youngest leaves. The presence of both the single-stranded and the double-stranded DNA forms was verified by Southern blots.

The same lack of symptoms was also observed in transgenic tobacco (*N. tabacum*, cv. *Xanthi*), derived from regenerated shoots following leaf disc transformation with LBA4404 pBin19/TYLCV-S 1.8 (31). Nevertheless these plants contained the same forms of viral DNA as the agroinoculated tomatoes (data not shown).

Bemisia tabaci transmits the disease from agroinoculated plants

In order to find out whether the single cloned DNA represents the complete TYLCV genome, whiteflies were fed on tomatoes diseased following agroinoculation. 12 out of 16 test plants that had been exposed to *B.tabaci* fed on agroinoculated plants exhibited clear symptoms of the TYLCV disease after three weeks. The presence of TYLCV-specific DNA was detectable in leaf squash blots already one week before the appearance of symptoms. Plants not showing any symptoms one month after exposure to *B.tabaci* also contained no detectable viral DNA. An example of the TYLCV DNA in the plants infected by *B.tabaci* fed on agroinoculated plants is shown in Figure 4.

The symptoms observed on *L.esculentum* cv. Marmande following transmission by *B.tabaci* were the same as those produced after grafting. A separate experiment comprising two lots of 15 plants each, grafted with either field infected material or agroinfected plants was carried out. Disease development was monitored over more than six weeks and special care was taken to insure identical growth conditions for the plants by repeatedly randomizing their positions in the glasshouse. No difference in severity and rate of development of symptoms between the grafted field material and the agroinoculated plants was observed.

Table 2. Pairwise comparisons of deduced amino-acid sequences encoded by *B.tabaci* transmitted geminiviruses

	V1	V2	C1	C2	C3	C4
1: TYLCV-S/TYLCV-I	83	86	77	67	64	50
2: TYLCV-S/ACMV	72	77	74	64	70	42
3: TYLCV-S/ICMV	70	75	76	52	55	60
4: TYLCV-S/TYLCV-T	66	70	75	57	62	49
5: TYLCV-I/ACMV	73	80	73	64	72	36
6: TYLCV-I/ICMV	72	79	75	62	66	56
7: TYLCV-I/TYLCV-T	72	73	78	58	67	63
8: TYLCV-T/ACMV	67	73	72	59	64	32
9: TYLCV-T/ICMV	74	79	77	63	63	44
10: ACMV/ICMV	77	77	74	60	63	45
11: TYLCV-S/ABMV	-	72	67	52	50	66
12: TYLCV-S/TGMV	-	72	67	60	51	62
13: TYLCV-S/BGM	-	70	63	53	50	59
14: TYLCV-S/SqLCV	-	70	53	56	46	28

Amino acid identity is given in percent derived with the program *BestFit* of the UWGCG-Sequence analysis package. The word size (K-tuple) was 2, the gap weight was 2.0 and gaplength weight was O.1. The program *FastA* yielded the the same relationship.

Sequences were from:

ABMV: Abutilon mosaic virus (54)
ACMV: African cassava mosaic virus (30)
BGMV: Bean golden mosaic virus (34)
ICMV: Indian cassava mosaic virus (D. J. Robinson and Y.G. Hong, pers. comm.)
SqLCV: Squash leaf curl virus (39)
TGMV: Tomato golden mosaic virus (33)
TYLCV-I: Tomato yellow leaf curl virus, Israeli isolate (35)
TYLCV-S: Tomato yellow leaf curl virus, Sardinian isolate (this study)
TYLCV-T: Tomato yellow leaf curl virus, Thai isolate (R. Beachy, pers. comm.)

TYLCV from Sardinia has no B-like DNA

In order to search for a potential B-like DNA that may have gone undetected clone banks derived from infected plants from both Sardinia and Sicily were screened with a probe specific for the intergenic region of TYLCV-S. In one particular experiment the DNA was digested with *Hind* III, an enzyme that cuts within the intergenic region and an additional digestion with *Sst* I was carried out to reduce the background of repeated cloning of the same TYLCV-S ('A') DNA. However, we did not find any hint for a B-DNA. Furthermore, using a PCR-based detection technique similar to the one described for the identification of geminiviruses in the grasses by Rybicki and Hughes (32), we have tried to amplify a potential B component in DNA prepared from naturally infected tomatoes. For this purpose we used one specific primer complementary to bases No. 2758 through 9 within the stemloop in the intergenic region and two different degenerate primers, based on the most conserved amino acid motifs in the BR1 and BL1 regions of African cassava mosaic virus (ACMV) (30), tomato golden mosaic virus (TGMV) (33) and bean golden mosaic virus (BGMV) (34). Again, no B-like sequences were amplified.

A 'geographical gradient' of similarity

The comparison of the DNA sequence of TYLCV-S with the DNA sequences of the two other isolates of TYLCV and the A DNAs of other whitefly-transmitted geminiviruses shows that TYLCV-S is closest to TYLCV-I (77% nucleotide identity), followed by African cassava mosaic virus (ACMV) (73%), Indian cassava mosaic virus (ICMV) (72%) and TYLCV-T (71%).

To see where the TYLCV isolate from Sardinia may be placed within the other whitefly-transmitted geminiviruses, we carried out a detailed pairwise comparison of the deduced protein sequences encoded by the genome of TYLCV-S with the corresponding sequences of the whitefly-transmitted geminiviruses currently available. The results of this comparison are given in Table 2.

When the amino acid sequences of the viral capsid proteins, encoded by ORF V2 of virion strand polarity serve as the basis for comparison, TYLCV-S is placed next to the TYLCV isolate from Israel (TYLCV-I, 86% amino acid identity), ACMV, Kenyan isolate, (77%) and ICMV, (75%). The TYLCV isolate from Thailand (TYLCV-T) is the most distant one to TYLCV-S (70%), whereas it ranges closest to Indian cassava mosaic virus (79%). The TYLCV isolate from the Near East (TYLCV-I) is about equidistant from both African (80%) and Indian cassava mosaic virus (79%); (Table 2). A very similar relationship becomes apparent when the deduced amino acid sequences of V1 are compared pairwise. A gradient of similarity among the amino acid sequences encoded by the ORFs of the virion sense strand of these geminiviruses ranging from the Sardinian TYLCV via the two cassava mosaic viruses to the Thai TYLCV reflects the geographical distribution of the respective viruses.

When the amino acid sequences of the proteins encoded by the ORFs of minus-strand polarity are compared, again the three TYLCV isolates do not form a homogeneous subgroup within the other whitefly-transmitted geminiviruses, but also no 'geographical gradient' of similarity is evident.

DISCUSSION

The genome organization of tomato yellow leaf curl virus from Sardinia (TYLCV-S), as reflected by the sequence of its 2770 nucleotides resembles the A DNA of the whitefly-transmitted geminiviruses described until now. Open reading frames are found on both strands of the DNA and transcription start signals are located in an intergenic region, equivalent to the common region of the genomes of other whitefly-transmitted geminiviruses. Preliminary S1-nuclease and primer extension mapping of transcripts confirm the bidirectional transcription mode of the TYLCV genome (Bendahmane *et al.*, in preparation).

From our results the following points are clear: First, the cloned molecule of 2770 base pairs, once transferred to tomato by *A.tumefaciens* Ti-plasmid vectors, is sufficient to cause a disease indistinguishable from that induced by natural infection. Second, the vector *B.tabaci* readily transmits the disease from plants agroinfected with this single molecule to healthy plants where typical symptoms result. Third, in these plants the viral DNA forms are the same as those in naturally infected plants. These results indicate that the full biological cycle can be reproduced experimentally using the cloned single genome of TYLCV-S.

Similar results have been recently obtained by Navot *et al.*, (35) for the TYLCV isolate from Israel.

Rochester *et al.*, (15) cloned a DNA A-like genome of a TYLCV isolate from Thailand (TYLCV-T) that is apparently accompanied by a DNA B, as is the case of all other whitefly-transmitted geminiviruses described. However, the presence of this B genome is not required for the elicitation of the disease symptoms, neither in *L.esculentum* or in *Nicotiana benthamiana*.

The degree of DNA sequence diversity between the three different isolates of TYLCV, 23% between TYLCV-S and TYLCV-I, 25% between TYLCV-I and TYLCV-T and 28%

between TYLCV-S and TYLCV-T, is in pronounced contrast to the usual very high sequence similarity between different geminivirus isolates or strains. For example, the DNA sequences of the Kenyan and Nigerian isolates of ACMV show 96% identity (36), the Kenyan, Nigerian and South African isolates of maize streak virus vary maximally by about 2% (37).

The remarkable difference between the (monopartite) TYLCV isolates from Sardinia and Israel and the bipartite one from Thailand immediately elicits the question of the origin and the relationship of these geminiviruses, in particular if they are compared with African and Indian cassava mosaic virus.

The comparison of the capsid protein (ORF V2) sequences reveals a peculiar 'geographical gradient of similarity' between TYLCV-S, TYLCV-I, ACMV, ICMV and TYLCV-T and prompts some speculation about its biological significance. The capsid protein of a geminivirus determines the specificity of insect transmission (38) and the capsid acts as an interface to the 'outside world'. Its interactions have to occur with factors specified by the insect vector, in this case the whitefly *B.tabaci*, as well as the natural plant host. Thus the relationship of the amino acid sequences of the capsid proteins reflects the impact of parameters that vary geographically, as for instance regional diversities in whitefly populations and/or factors specified by different wild host plants, rather than the host cultivated tomato. Note, that for instance the capsid protein sequence of TYLCV-S shows only 72% identity (Table 2) with that of tomato golden mosaic virus (TGMV), another geminivirus of the same host tomato. This is comparable with its similarity to bean golden mosaic virus (BGMV) (70%) or squash leaf curl virus (SqCLV), (39), (70%), geminiviruses of different host plants. In this context it might be interesting to determine the relationship with tomato leaf curl virus (TomLCV) from India which, based upon dot hybridisation data is claimed be closer to ACMV than ICMV (40).

A reason for the same gradient of relative similarity between the V1 sequences as it occurs between the V2 sequences is not immediately apparent. The exact function of the V1 protein is not clear to date. It is necessary for the elicitation of symptoms among the leafhopper-transmitted geminiviruses, probably because it is required for an efficient systemic spread of the virus throughout the host plant (41,42). For viral DNA replication, however, its expression is not required, neither in leafhopper- nor whitefly-transmitted geminiviruses (26,41). It also is dispensable for the systemic spread of ACMV (43) and an ORF V1 does not exist in the geminiviruses classified as 'New World geminiviruses' by Howarth and Vandemark, (44). However, in TYLCV-S the introduction of two premature stop codons at different positions into ORF V1 suppresses the disease symptoms and strongly reduces the movement of the viral DNA within the plant (Gronenborn *et al.*, in preparation). This proves the necessity of the V1 protein expression and is in line with TYLCV-S being a monocomponent geminivirus. In order to accomplish this movement, the V1 protein may have to interact with the viral capsid protein, and this may be the reason for the 'concerted evolution' of the two protein sequences.

'Similarity trees' displaying the relationship between the whitefly-transmitted geminiviruses as reflected by the amino acid sequences of V1 and V2 using the program PileUp of the UWGCG sequence analysis package (45) or the program CLUSTAL (PC-gene sequence analysis package, Intelligenetics) yielded the same relationship between the three isolates of TYLCV and African and Indian cassava mosaic virus, as is apparent from their pairwise comparisons. They constitute a subgroup of Old World geminiviruses as opposed to all other whitefly-transmitted geminiviruses (New World) as has been proposed by Howarth and Vandemark (44).

The pairwise comparison (Table 2) as well as similarity trees of the protein sequences encoded by the ORFs of the complementary strand, C1 through C4, does not group the TYLCVs together, but also their relationship is not the same as for the V1 and V2 proteins. Clearly, other parameters than the ones influencing V1 and V2 have an impact on the evolution of these protein sequences.

The C1 proteins are the only viral encoded proteins required for the replication of the geminivirus genome (7,46). The C2 proteins influence the ratio between double-stranded and single-stranded DNA, most probably by transactivation of the capsid protein expression and thus removing the single-stranded DNA from the equilibrium by encapsidation (47,48,49). The sequences of C3 of both TYLCV-S and TYLCV-I are significantly more similar to C3 of ACMV than to any other geminivirus (Table 2, lines 2 and 5), a reason for this remains to be determined. C4 is a small ORF entirely overlapped by ORF C1. It has been shown to be non essential for the bipartite geminiviruses TGMV and ACMV (7,50). However, it may have a role in TYLCV since this whitefly-transmitted geminivirus is infectious with one DNA genome.

The three different isolates of TYLCV represent an interesting example in the evolution of geminiviruses. A similar syndrome of disease is caused by three quite distant viruses, one of which (TYLCV-T) appears to have acquired a second genome component. The B-DNA associated with TYLCV-T seems to modulate symptom severity (15). In that sense it may be comparable to a sort of satellite. Similar effects on symptomatology have been described for the B-DNA of ACMV and deletion mutants thereof, as well as for TGMV (51,52).

After the first DNA sequence of a geminivirus (ACMV) was published in 1983 (30), Kikuno *et al.* (53) speculated, based upon the sequence comparison between DNA 1 and 2 of ACMV, that geminivirus genomes evolved from monopartite to bipartite via duplication. The three different isolates of TYLCV may represent actual examples within this process.

ACKNOWLEDGEMENTS

We are indebted to H. Czosnek for sharing cloned TYLCV-I DNA, sequence information and many stimulating discussions. We thank R. Beachy and D.J. Robinson for sequence data prior to publication and J.C. Mercier and H. Laterrot for tomato seeds. The advice and help of P. Caciagli in handling *B.tabaci* as well as the excellent technical assistence of F. de Kouchkovsky and F. Jouanneau are gratefully acknowledged. We thank J.M. Bonneville for critical comments on and C. Deforeit for help with the manuscript and R.Esnault and J.C. Barbet for help with computer work. G.P.A was in receipt of an EMBO fellowship; part of this work was supported by a grant of the *Bundesminister für Forschung und Technologie* to B.G.

REFERENCES

1. Goodman, R.M. (1981) In Kurstak, E. (Ed.), Handbook of Plant Virus Infections and Comparative Diagnosis, Elsevier/North Holland, Amsterdam. pp. 883–910.
2. Harrison, B.D., Barker, H., Bock, K.R., Guthrie, E.J., Meredith, G. and Atkinson, M. (1977) *Nature* **270**, 760–762.
3. Stanley, J. (1985) *Adv. Virus Res.* **30**, 139–177.

4. Davies, J.W. and Stanley, J. (1989).*Trends Genetics* **5**, 77–81.
5. Rogers, S.G., Bisaro, D.M., Horsch, R.B., Fraley, R.T., Hoffmann, N.L., Brand, L., Elmer, J.S. and Lloyd, A.M. (1986). *Cell,* **45**, 593–600.
6. Townsend, R., Watts, J. and Stanley, J. (1986) *Nucl. Acids Res.* **14**, 1253–1265.
7. Elmer, J. S., Brand, L., Sunter, G., Gardiner, W. E., Bisaro, D. M.and Rogers, S. G. (1988) *Nucl. Acids Res.* **16**, 7043–7060.
8. Cohen, S. and Harpaz I. (1964) *Entomol. Exp. Appl.* **7**, 155–166.
9. Cohen, S. and Nitzany, F.E. (1966) *Phytopathology* **56**, 1127–1131
10. Czosnek, H., Ber, R., Antignus, Y., Cohen, S., Navot, N. and Zamir, D. (1988) *Phytopathology* **78**, 508–512.
11. Czosnek, H., Navot, N. and Laterrot, H. (1990). *Phytopath. Medit.* **26**, 1–6.
12. Brown, J.K. and Nelson, M.R. (1988) *Plant Dis.* **72**, 866–869.
13. Luisoni, E., Milne, R.G., Caciagli, P., Accotto, G.P., Conti,M., Gallitelli, D., Martelli, G.P., Antignus, Y., Cohen, S., Marras, F., Idini, G., Loche, P., Piras, S. and Leoni, S. (1989). 6th Conference, Recent Advances in Vegetable Virus Research, August 27–31, Asilomar, California. pp. 13–14.
14. Credi, R., Butti, L. and Canova, A. (1989) *Phytopath. Medit.* **28**, 223–226.
15. Rochester, D.E., Kositrana, W. and Beachy, R.N. (1990) *Virology* **178**, 520–526.
16. Navot, N., Ber, R. and Czosnek, H. (1989) *Phytopathology* **79**, 562–568.
17. Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
18. Vieira, J. and Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
19. Sambrook J., Fritsch, E.F. and Maniatis, T.(1989) Molecular cloning, a laboratory manual, Cold Spring Harbor Laboratory Press.
20. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
21. Grimsley, N. and Bisaro, D. (1987) In Hohn, T. and Schell, J., (eds) *Plant DNA Infectious Agents*, Springer, Wien, New York. pp. 87–107
22. Bevan, M. (1984) *Nucleic Acids Res.* **12**, 8711–8721.
23. Höfgen, R. and Willmitzer, L. (1988) *Nucleic Acids Res.* **16**, 9877.
24. Hoekema, A., Hirsch, P.R., Hooykaas, P.J.J. and Schilperoort, R.A. (1983) *Nature* **303**, 170–180.
25. Ooms, G., Hooykaas, P.J.J., Van Veen, H.J.M., Van Beelen, P., Regensburg-Tuink, T.J.G. and Schilperoort, R.A. (1982) *Plasmid* **7**, 15–29.
26. Matzeit, V., Schaefer, S., Schalk, H.-J., Kammann, M., Schell, J. and Gronenborn, B. (1991) *Plant Cell* **3**, 247–258.
27. Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
28. Lazarowitz, S.G. (1987) *Plant Mol. Biol. Reporter* **4**, 177–192.
29. Stanley, J., Markham, P.G., Callis, R.J. and Pinner, M.S. (1986). *EMBO J.* **5**, 1761–1767.
30. Stanley, J. and Gay, M.R. (1983) *Nature* **301**, 260–262.
31. Horsch, R.B., Fry, J.E., Hoffmann, N.L., Wallroth, M., Eichholtz, D., Rogers, S.G. and Fraley, R.T. (1985) *Science* **227**, 1229–1231.
32. Rybicki, E.P.and Hughes, F.L. (1990) *J. Gen. Virol.* **71**, 2519–2526.
33. Hamilton, W.D.O., Stein, V.E., Coutts, R.H.A. and Buck, K.W. (1984) *EMBO J.* **3**, 2197–2205.
34. Howarth, A.J., Caton, J., Bossert, M. and Goodman, R.M. (1985) *Proc. Natl. Acad. USA* **82**, 3572–3576.
35. Navot, N., Pichersky, E., Zeidan, M., Zamir, D. and Czosnek, H. (1991) *Virology*, **185**, 151–161.
36. Morris, B., Coates, L., Lowe, S., Richardson, K. and Eddy, P. (1990) *Nucl. Acids Res.* **18**, 197.
37. Lazarowitz, S.G. (1988) *Nucl. Acids Res.* **16**, 229–249.
38. Briddon, R.W., Pinner, M.S., Stanley, J. and Markham, P.G. (1990). *Virology*, **177**, 85–94.
39. Lazarowitz, S.G. and Lazdins, I.B. (1991) *Virology* **180**, 58–69.
40. Harrison, B.D., Muniyappa, V., Swanson, M.M., Roberts, I.M. and Robinson, D.J. (1991) *Ann. appl.* Biol. **118**, 299–308.
41. Boulton, M.I., Steinkellner, H., Donson, J., Markham, P.J., King, D.I. and Davies, J.W. (1989) *J. Gen. Virol.* **70**, 2309–2323.
42. Lazarowitz, S. G., Pinder, A. J., Damsteegt, V. D. and Rogers, S. G. (1989) *EMBO J.* **8**, 1023–1032.
43. Etessami, P., Watts, J. and Stanley, J. (1989) *J. Gen. Virol.* **70**, 277–289.
44. Howarth, A.J. and Vandemark, G.J. (1989) *J. Gen. Virol.* **70**, 2717–2727.
45. Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucl. Acids Res.* **12**, 387–395.
46. Hanley-Bowdoin, L., Elmer, J. S. and Rogers, S. G. (1990) *Proc. Natl. Acad. Sci. USA* **87**,1446–1450.
47. Hayes, R.J. and Buck, K.W. (1989) *Nucl. Acids Res.* **17**, 10213–10222.
48. Sunter, G., Hartitz, M.D., Hormuzdi, S.G., Brough, C.L. and Bisaro, D.M. (1990) *Virology* **179**, 69–77.
49. Sunter, G. and Bisaro, D.M. (1991) *Virology* **180**, 416–419.
50. Etessami, P., Sanders, K., Watts, J. and Stanley, J. (1991) *J. Gen. Virol.* **72**, 1005–1012.
51. Revington, G.N., Sunter, G. and Bisaro, D.M. (1989) *Plant Cell* **1**, 985–992.
52. Stanley, J., Frischmuth, T. and Ellwood, S. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 6291–6295.
53. Kikuno, R., Toh, H., Hayashida, H. and Miyata, T. (1984) *Nature* **308**, 562.
54. Frischmuth, T., Zimmat, G.and Jeske, H. (1990) *Virology* **178**, 461–468.